

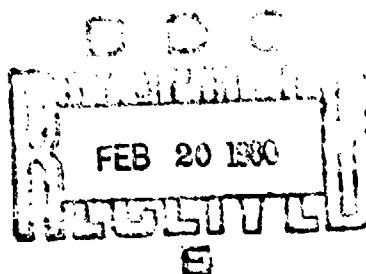
LEVEL 2

NPS 54-79-006

# NAVAL POSTGRADUATE SCHOOL

## Monterey, California

DDC FILE COPY  
ADA 080940



NAVAL OFFICER RETENTION AS A FUNCTION OF COMMISSION

SOURCE AND FIRST AND SECOND DUTY ASSIGNMENTS:

AN EVALUATION OF THREE ESTIMATION MODELS

R. A. Weitzman

David W. Robertson

September 1979

Approved for public release; distribution unlimited.

Prepared for:

Navy Personnel Research and Development Center  
San Diego, CA 92152

80 2 19 187

NAVAL POSTGRADUATE SCHOOL  
Monterey, California

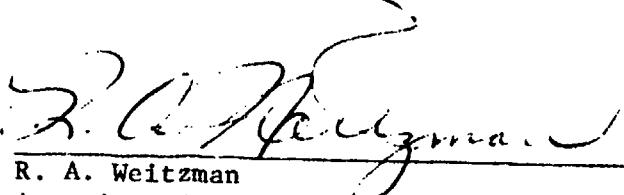
Rear Admiral T. F. Dedman  
Superintendent

Jack R. Borsting  
Provost

This work was supported by the Navy Personnel Research and Development Center, San Diego, California, under its acquisition and initial service programs.

Reproduction of all or part of this report is authorized.

This report was prepared by:

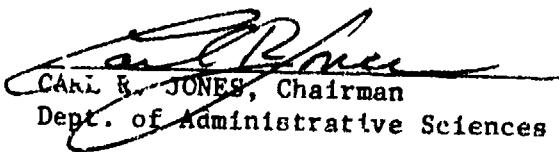
  
R. A. Weitzman

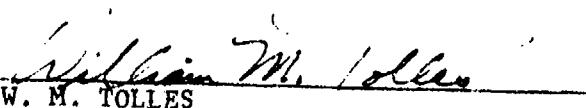
Associate Professor

  
David W. Robertson

Navy Personnel Research and Development  
Center

Reviewed by:

  
CARL R. JONES, Chairman  
Dept. of Administrative Sciences

  
W. M. TOLLES  
Dean of Research

## UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS 54-79-006	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NAVAL OFFICER RETENTION AS A FUNCTION OF COMMISSION SOURCE AND FIRST AND SECOND DUTY ASSIGNMENTS: AN EVALUATION OF THREE ESTIMATION MODELS		5. DATE OF REPORT & PERIOD COVERED Final Report Sep 1976 - Nov 1977
6. AUTHOR(S) R. A. Weitzman David W. Robertson		7. PERFORMING ORGANIZATION NUMBER NPS 54-79-006
8. CONTRACT OR GRANT NUMBER(S)		
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63707N 20107-PN.02A
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, CA 92152		12. REPORT DATE September 1979
13. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12/33		14. NUMBER OF PAGES 33
15. SECURITY CLASS. (of this report) UNCLASSIFIED		16. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. 16 20107PN		
18. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) 17 20107PN		
19. SUPPLEMENTARY NOTES		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In an assignment system characterized by a variety of individuals and jobs, an important problem is how to assign the most appropriate individual to each job. Linear programming algorithms have been found useful for optimal assignment of individuals. The objective function of interest (e.g., retention) is optimized in these algorithms from data in a "cost" matrix (e.g., a matrix of retention proportions). The optimization achieved is dependent in part upon the stability of the "cost" data. When the "cost" data are retention proportions for all possible assignment patterns in a		

251450

source-to-assignment matrix of several hundred possible patterns, the observed retention proportions may be very unstable for patterns with few individuals, or even unavailable for patterns with no individuals. In this study, configural (or pattern) analysis models were evaluated for accuracy and stability in providing estimates of retention proportions for a source-to-assignment matrix with several hundred possible assignment patterns.

The three Structural Pattern Analysis (SPA) models developed and evaluated were True-score, Linear-covariance, and Independence. The dichotomous criterion variable was the retention outcome of Navy officer personnel who had completed their initial service obligation. The three predictor variables were training source (5 sources), first assignment (6 categories), and second assignment (6 categories). These variables thus provided a framework for a source-to-assignment matrix of  $5 \times 6 \times 6 = 180$  patterns (cells). Results of analyses involving a different, 8-category assignment-classification system, also reported, tend to corroborate the 6-category results.

The major finding was that one of the SPA models provided more stable data than did the calculations based on the actual outcomes. This finding suggests that stable estimates of personnel retention proportions are possible for use with a source-to-assignment matrix in algorithms for optimizing the assignment of personnel.

UNCLASSIFIED

## FOREWORD

This study was conducted in response to Navy Decision Coordinating Paper, Personnel Supply Systems (NDCP-Z0107-FN), under subproject PN.02A, Career Officer Retention, and under the sponsorship of the Deputy Chief of Naval Operations for Manpower (OP-01). The overall objectives of the subproject are to develop career paths that enable junior officers to make long-term career plans and to assist the Navy in developing assignment strategies that increase career retention of quality Naval officers.

The study was undertaken to identify patterns in the duty assignment system that are associated with retention. If patterns are identified that are controllable through the assignment system, alternative strategies that increase retention may be developed.

In a prior study (Robertson & Pass, 1979), the association of the type of first assignment with retention was demonstrated. The present study addresses a technical problem concerned with the instability of small sample sizes. Analysis of assignment sequences (patterns) shows that the frequency of alternative patterns increases exponentially with the number of available assignment types, resulting in small, unstable samples. Configural-analysis models may provide data more stable than raw data for the testing of alternative assignment strategies.

The substantial and valuable assistance of the following persons is gratefully acknowledged: Pat Meadows for programming and data processing, John Pass for data processing, and Hazel F. Schwab and Montez Bunten for clerical support.

R. A. WEITZMAN  
D. W. ROBERTSON

to Form 1473 (back page)

Accession For	
NTIS GRAAI	
DDC TAB	
Unclassified	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Available and/or special
A	

## SUMMARY

### Problem

In an assignment system characterized by many different individuals and jobs, an important problem is how to assign individuals to jobs most appropriately. Linear programming algorithms have been found useful in solving this problem. Using test scores and other individual data as predictors, these algorithms optimize a criterion of interest, like personnel retention. The optimization achieved is dependent for accuracy, however, on the stability of the criterion data. When these data are retention proportions for all possible patterns of predictor-variable values in a matrix of several hundred possible patterns, the retention proportions may be not only very unstable for patterns containing few individuals but even unavailable for patterns containing no individuals.

### Purpose

Configural (or pattern) analysis models were evaluated for accuracy and stability in providing estimates of retention proportions for a source-to-assignment matrix of several hundred possible patterns. The data were the early assignment patterns and retention outcomes of Navy officers (Unrestricted Line designator) from five Commission Sources. The tasks specifically addressed were to (1) develop three Structural Pattern Analysis (SPA) models--True-score, Linear-covariance, and Independence; (2) estimate, by the SPA models, the dichotomous criterion of retention, given three variables--Commission Source, Initial Duty Assignment, and Second Duty Assignment; and (3) cross-validate the estimates, particularly for assignment patterns of small sample size.

### Approach

From an inventory of several hundred possible assignments for Navy officers, a small number of assignment categories were constructed--6 Ship-type categories in one classification system, and 8 Retention-probability categories in another. The three predictor variables were (1) Commission Source (5 sources), (2) First Assignment (6 or 8 categories), and (3) Second Assignment (6 or 8 categories). Thus, one source-to-assignment matrix analyzed contained  $5 \times 6 \times 6 = 180$  patterns (or cells), and the other 320 cells. Each cell was randomly divided to provide a double cross-validation design. Cells with the largest and smallest sample sizes were analyzed separately. The SPA retention-probability estimates from one subgroup were double cross-validated with the observed (actual) retention proportions of the other subgroup to measure accuracy (or validity). The Observed-Observed and Estimated-Estimated correlations for the two subgroups were used as measures of stability (or reliability).

### Findings

The Independence model was both the most accurate and the most stable of the three SPA models evaluated, and it also provided more stable data than did the calculations based on the actual outcomes.

### Conclusions

Structural Pattern Analysis (SPA) models can provide stable estimates of personnel-retention proportions for possible use with linear-programming algorithms in a source-to-assignment matrix to minimize personnel losses. The Independence model does particularly well for matrices having cells that contain few or no individuals.

## CONTENTS

	Page
INTRODUCTION . . . . .	1
Problem . . . . .	1
Background. . . . .	1
Purpose . . . . .	2
APPROACH . . . . .	3
Sample. . . . .	3
Assignment Categories . . . . .	3
SPA Models. . . . .	5
Analysis. . . . .	5
Retention Proportions. . . . .	5
Validation . . . . .	6
RESULTS. . . . .	6
DISCUSSION . . . . .	10
CONCLUSIONS. . . . .	11
REFERENCES . . . . .	12
APPENDIX A--OFFICER ASSIGNMENT CATEGORIES BY UNIT-TYPE . . . . .	A-0
APPENDIX B--DESCRIPTION OF STRUCTURAL PATTERN ANALYSIS MODELS. .	B-0
APPENDIX C--RETENTION ESTIMATES BY THE SPA INDEPENDENCE MODEL. .	C-0

### DISTRIBUTION LIST

#### LIST OF TABLES

1. Assignment Categories for Classification Systems . . . . .	4
2. Relationship of SPA-model Estimates to Observed Retention Proportions . . . . .	7
3. Double Cross-validation of Retention Proportions Estimated by SPA Models. . . . .	8

## INTRODUCTION

### Problem

If there is a relationship between the early duty-assignment patterns of Navy officers and retention, assignment strategies can be identified that increase retention or permit allocation of the best performing officers to high-retention paths. In a study of officers with the Unrestricted Line designator who were assigned to surface ships or shore installations for their first assignment from five commission sources, it was found that both the type of first assignment and the college education major, as well as the commission source itself, were associated with retention (Robertson & Pass, 1979). (Data were not available on other variables that may affect the assignment decision, e.g., officer class standing and officer assignment preference.) Of the great number and variety of jobs that must be filled in performing Navy missions, some may provide better opportunities for career enhancement and motivation than others. Since all of the jobs are considered essential to carry out the various missions, it is not feasible to minimize assignments to low-retention jobs and maximize assignments to the others. However, it would be reasonable to try to increase retention by determining the retention outcomes for various assignment patterns from the present allocation procedure and to use this information in future officer allocation.

A particular difficulty in evaluating alternative allocation strategies stems from the instability of the obtained retention proportions for source-to-assignment patterns containing few or no officers. This instability reduces the accuracy of linear-programming algorithms that, with stable "cost" or benefit data (e.g., retention proportions or test scores), have been found useful in providing optimal "transportation" of individuals from origins to destinations (Robertson & Montague, 1976). To increase officer retention by the use of optimal source-to-assignment strategies, therefore, stable as well as accurate estimates of retention for all source-to-assignment patterns are necessary. Since the predictors in this problem are categorical (e.g., Commission Source), configural (or pattern) analysis models may be useful in providing the estimates.

### Background

In one approach to using polychotomous item responses to predict performance on a continuous criterion, each individual is assigned the mean criterion measurement or score of all individuals who have the same item response pattern (Meehl, 1950; Gaier & Lee, 1953; Lubin & Osborn, 1957; Lykken & Rose, 1963; Horst, 1968; Weitzman, 1973a). If the criterion is income, for example, the mean income of male college graduates is the predicted criterion score of an individual who responds on a questionnaire that he is a male and that he is a college graduate. The criterion itself may be polychotomous, but in this case the predicted criterion score for each item-response pattern depends on the criterion category and is equal to the proportion of individuals having the pattern who are in the criterion category (Lubin & Osborn, 1960). In the particular case of a dichotomous criterion consisting of the two values, 1 for success and 0 for failure,

the mean criterion score for a pattern is the proportion of individuals having the pattern who have the value of 1 on the criterion, and this proportion is interpretable as the probability of success (Weitzman, 1973b).

A sizable ratio of individuals to items is required if the number of individuals having each response pattern is to be large enough to make the pattern scores reliable. The size of this ratio depends on the number of classifiable responses to each item. If for every item this number is two (correct/incorrect or yes/no) and if reliability requires a mean of 20 individuals per response pattern, then for  $K$  items there are  $2^K$  possible response patterns and the total number of individuals must be  $20(2^K)$ . For  $K = 5$ , this number is  $20(32)$ , or 640, implying an individuals-to-items ratio of 640-to-5, or 128-to-1.

Even a mean of 20 individuals per response pattern may not be sufficient, however, if the variation from response pattern to response pattern is large. If this is the case (as it tends to be in the present assignment data), there may be a number of response patterns for which pattern scores are indeterminable because no one has them. A major practical problem of pattern analysis is the occurrence of vacant or sparsely populated response patterns.

This problem is solvable for polychotomous criteria if the observed distribution of frequencies over response patterns is an approximation of a theoretical distribution. If the individuals observed constitute a sample from a population, for example, an observed zero frequency may be an estimate of a true non-zero frequency.

#### Purpose

This report evaluates three models for the estimation of the proportions for cells in a source-to-assignment matrix when cell sample sizes are too small for direct calculation of cell proportions. As applied to the data of the present study (proportions for all patterns of officer commission sources and initial duty assignments), the tasks specifically addressed were these:

1. Develop three Structural Pattern Analysis (SPA) models--the True-score, Linear-covariance, and Independence models.
2. Estimate the dichotomous criterion of retention, given three variables--Commission Source, Initial Duty Assignment, and Second Duty Assignment--defining each pattern.
3. Cross-validate the estimates, particularly for patterns of small sample size.

## APPROACH

### Sample

Officers with the Unrestricted Line designator (11XX) whose Active Commission Base Date (ACBD) was within the years 1966 through 1970 formed the population studied. The officers who were still on active duty at least 2 years beyond their initial Minimum Service Requirement (MSR) were identified as "career." These data were the most current available for a stable retention criterion. The officers (total N = 7616) were from one of the following Commission Sources:

1. Naval Academy (ACAD)--5 years Minimum Service Requirement (MSR) incurred.
2. Naval Reserve Officers Training Corps-Scholarship (NROTC-SCL)--4 years MSR incurred.
3. Naval Reserve Officers Training Corps-College (NROTC-COL)--3 years MSR incurred.
4. Officer Candidate School (OCS)--3 years MSR incurred.
5. Reserve Officer Candidate (ROC)--3 years MSR incurred.

The record of each officer's initial and second duty assignment was reconstructed from data on the Officer Master Tape maintained by the Bureau of Naval Personnel. The sample selected for analysis was not representative of the population because only those officers were sampled who were transferred to a second assignment (about half of the actual population) prior to completing the MSR for their particular Commission Source. The primary purpose of this sampling procedure was to permit testing of the analytical models with first- and second-assignment data for every member of the sample.

### Assignment Categories and Study Variables

Patterns of first and second assignments were created for two different systems of assignment classification: a Ship-type system of six categories and a Retention-probability system of eight categories (see Table 1). The categories in both systems are composites built from the 43 Unit-type categories developed by Robertson & Pass (1979) from the several hundred Ship and Station Codes of the Officer Classification Manual (NAVPER 15839C Vol. I). (Table 2 of the Robertson-Pass study is reproduced here as Appendix A.) Examination of Table 1 is sufficient to make clear the formation of Ship-type categories, but the formation of Retention-probability categories requires some explanation. The Unit-types contained in a Retention-probability category all have approximately equal retention probabilities that tend to differ from the retention probabilities of Unit-types contained in other Retention-probability categories. The Robertson-Pass study provides the Unit-type retention probabilities used to form the Retention-probability categories.

Table 1  
Assignment Categories for Classification Systems

Category Title	Unit-Type Source <sup>a</sup>
Ship-type Categories	
1. Primary Combatant Ship--Small	10, 11, 26, 8
2. Primary Combatant Ship--Large	2, 7, 6, 3
3. Combat Support Ship	18, 9, 5, 4, 27
4. Logistic Support Ship	29, 24, 16, 15, 25, 14, 17, 13
5. Fleet/Joint/Allied Sqd Staff	1, 19, 41, 12, 42
6. Shore	22, 37, 40, 33, 36, 39, 30, 32, 23, 31, 35, 43, 34, 20, 33, 28, 21
Retention-probability Categories	
1. Fleet	10, 11, 26
2. Fleet	8
3. Fleet	2, 18, 9, 29, 7
4. Fleet--Amphibious	5, 4
5. Fleet	24, 16, 15, 6, 27
6. Fleet	3, 25, 14, 17, 13
7. Fleet--Staff	1, 19, 41, 12, 42
8. Shore	22, 37, 40, 38, 36, 39, 30, 32, 23, 31, 35, 43, 34, 20, 33, 28, 21

<sup>a</sup>See Appendix A for titles of Unit-types.

The three predictor variables, with their number of categories, are indicated below.

<u>Variable</u>	<u>Number of Categories</u>
$X_1$ Commission Source	5
$X_2$ First Assignment	6 or 8 (Table 1)
$X_3$ Second Assignment	6 or 8 (Table 1)

Thus, the source-to-assignment matrix created with the use of Ship-type categories contained  $5 \times 6 \times 6 = 180$  cells, and the matrix created with the use of Retention-probability categories contained 320 cells. The criterion variable was the dichotomous retention status, career (1) or non-career (0).

#### Estimation Models

Preliminary work developed and evaluated a number of different models for the estimation of proportions of individuals within patterns. The three most promising of these models were chosen for investigation in this study: (1) the True-score model, (2) the Linear-covariance model, and (3) the Independence model. Appendix B provides a technical description of these three models.

#### Analysis

Double cross-validation of retention proportions, determined from both observed and estimated cell proportions, was used to evaluate the three estimation models.

The three models were evaluated on both the 180-cell matrix constructed from the 6 Ship-type categories and the 320-cell matrix constructed from the 8 Retention-probability categories. For each matrix, the data of each cell were randomly divided into Subgroups 1 and 2 (for the cross-validation) so that each cell's two subgroup sizes differed by no more than one individual.

#### Retention Proportions

The overall retention proportion for the total sample ( $N = 7616$ ) was .204. Analogous to this is the retention proportion for each pattern (cell) defined by a specific commission source and combination of first and second assignments.

For each subgroup of each cell, retention proportions were calculated both from the observed retention frequencies and from the model-estimated frequencies (see Appendix B). Thus, four sets of retention proportions were generated--Observed and Estimated for each subgroup.

### Validation

With the assignment patterns (cells) serving as "subjects" and the cell retention proportions as "scores," Pearson product-moment correlations ( $r$ ) were calculated from Observed (O) and Estimated (E) retention proportions both within and between Subgroups 1 and 2. Thus, the correlation coefficients below were calculated for each model.

<u>Validation</u>		<u>Double Cross-Validation</u>	
<u>Subgroup 1</u>	<u>Subgroup 2</u>	<u>Subgroup 2</u>	
$r_{O_1 E_1}$	$r_{O_2 E_2}$	$O_1$	$O_2$
		$E_1$	$E_2$

The rationale for evaluating the stability of the retention proportions is as follows: If the Estimated (E) proportions are more stable than the Observed (O) proportions, the values of  $r_{E_1 E_2}$  should be greater than  $r_{O_1 O_2}$ .

A similar rationale applies for evaluating the accuracy (validity) of the Estimated (E) retention proportions: The values of  $r_{E_1 O_2}$  and  $r_{O_1 E_2}$  should be largest for the most accurate of the three models and smallest for the least accurate of them.

Since estimates for small or zero-N cells were of particular interest, the largest and smallest cells of each matrix (i.e., largest 90 cells and smallest 90 cells of the 180-cell matrix) were analyzed separately, and zero-N cells were excluded from the calculation of all correlations. (Thus, fewer than the smallest 90 and smallest 160 cells were used for the calculations. Otherwise, the E-E correlations would have been based on more cells than the O-E, E-O, or O-O correlations, to which the zero-N cells could not contribute.)

### RESULTS

As shown in Table 2, the correlation between the observed retention proportions and the retention proportions estimated by each of the models in the same subgroup was highest for the True-score model and lowest for the Linear-covariance model. The high value for the True-score model was to be expected because the cell-proportion estimates yielded by this model are linear functions of the observed proportions. However, the correlations for the Linear-covariance model, which were lower than those for the

Independence model, were a surprise. The Linear-covariance model, which allows for possible non-zero covariances between predictor variables, ought to provide better estimates than the Independence model, which does not make such an allowance. (A possible explanation for these results is presented in the Discussion section.)

Table 2  
Relationship of SPA-model Estimates  
to Observed Retention Proportions

SPA model	Correlation	
	Subgroup 1 <sup>a</sup> $r_{O_1 E_1}$	Subgroup 2 <sup>a</sup> $r_{O_2 E_2}$
True-score	1.00	1.00
Linear-covariance	.57	.53
Independence	.64	.68

<sup>a</sup>For Subgroup 1, correlations were calculated on 157 of the 180 cells of the Ship-type matrix. The other 23 cells, which had zero assigned officers, were excluded. For Subgroup 2, 154 cells were used.

In the separate analyses of large and small cells in the double cross-validation design, the major finding was that the Independence model provided the best estimates in the stability ( $r_{EE}$ ) test and that  $r_{EE} > r_{OO}$ . (For the Ship-type categories of the 180-cell matrix, large cells averaged about 17 officers, and small cells about 5.) Table 3 presents the results of these analyses.

Comparison of correlations for the Observed-Observed, Observed-Estimated, and Estimated-Estimated relationships between the two subgroups shows that there are no differences for the True-score model--all correlations are about .84 for the large cells and .61 for the small cells of the Ship-type (180-cell) matrix, and about .68 for the large and .41 for the small cells of the Retention-probability (320-cell) matrix (see Table 3). This result again reflects the fact that the model's estimates are linear functions of the observed values.

Table 3  
Double Cross-Validation of Retention Proportions  
Estimated by SPA Models

Model	Correlation <sup>a</sup>			
	Largest cells <sup>b</sup>		Smallest cells <sup>c</sup>	
	Observed $O_2$	Estimated $E_2$	Observed $O_2$	Estimated $E_2$
Ship-type Categories				
True-score	$O_1$	.84	.85	.61
	$E_1$	.84	.85	.61
Linear-covariance	$O_1$	.84	.84	.61
	$E_1$	.83	.95	.38
Independence	$O_1$	.84	.88	.61
	$E_1$	.89	.99	.60
Retention-probability Categories				
True-score	$O_1$	.67	.67	.42
	$E_1$	.68	.69	.42
Linear-covariance	$O_1$	.67	.73	.42
	$E_1$	.73	.94	.19
Independence	$O_1$	.67	.81	.42
	$E_1$	.81	.99	.51

<sup>a</sup>Subscripts identify Subgroups 1 and 2 (e.g.,  $r_{E_1 O_2}$  is the relationship between the estimated retention proportions of Subgroup 1 and the observed retention proportions of Subgroup 2).

<sup>b</sup>Cells with zero officers assigned were excluded from the correlations. For the Ship-type categories,  $N = 90$  cells for each subgroup; for the Retention-probability categories,  $N = 160$  for each subgroup.

<sup>c</sup>For the Ship-type categories,  $N = 67$  cells for Subgroup 1 and 64 cells for Subgroup 2; for the Retention-probability categories,  $N = 111$  cells for Subgroup 1 and 117 cells for Subgroup 2.

For the Linear-covariance model (see Table 3), in the case of the large cells, the two Observed-Estimated correlations are about equal to or slightly larger than the Observed-Observed correlation ( $r_{OE} \approx r_{EO} \approx r_{OO} \approx .84$  for the Ship-type categories, and  $r_{OE} = r_{EO} = .73$  and  $r_{OO} = .67$  for the Retention-probability categories) whereas, in the case of the small cells, the two Observed-Estimated correlations are substantially lower than the Observed-Observed correlation ( $r_{OE} = .19$ ,  $r_{EO} = .38$ , and  $r_{OO} = .61$  for the Ship-type categories, and  $r_{OE} = .26$ ,  $r_{EO} = .19$ , and  $r_{OO} = .42$  for the Retention-probability categories). The Estimated-Estimated large-to-small-cell drop for this model (from .95 to .70 for the Ship-type categories and from .91 to .69 for the Retention-probability categories) appears to reflect a large error component in the model's estimates for the small cells.

For the Independence model (again, see Table 3), the two Observed-Estimated correlations are larger than the Observed-Observed correlation in the case of the large cells ( $r_{OE} = .88$ ,  $r_{EO} = .89$ , and  $r_{OO} = .84$  for the Ship-type categories, and  $r_{OE} = r_{EO} = .81$  and  $r_{OO} = .67$  for the Retention-probability categories) and about equal or larger in the case of the small cells ( $r_{OE} \approx r_{EO} \approx r_{OO} \approx .60$  for the Ship-type categories, and  $r_{OE} = .39$ ,  $r_{EO} = .51$ , and  $r_{OO} = .42$  for the Retention-probability categories). The Independence model is also the only one that demonstrates a highly stable Estimated-Estimated relationship for both large and small cells ( $r = .99$  for both the Ship-type and Retention-probability categories).

Since the Independence model appears to be the most useful for generating Retention-probability estimates, particularly for patterns containing few or no individuals, a sample of the output of this model is displayed in Appendix C for some high- and low-retention Ship-type patterns (Tables C-1 and C-2) and some Retention-probability patterns (Tables C-3 and C-4). Pattern 141 in Table C-3 provides an interesting example of the stability of the estimates for small cells by the Independence model. In this pattern, with cell size  $N = 2$  in each of the two subgroups, the observed proportions are .50 and 1.00 respectively, but the estimates are very similar--.82 and .79. Pattern 133 (of Table C-3), with cell sizes of  $N = 3$  each, also demonstrates a similar large difference between the observed proportions and a small difference between the estimates.

## DISCUSSION

If the Observed-Observed and Estimated-Estimated relationships are conceptualized as measures of stability or reliability, and the Observed-Estimated relationships as measures of accuracy or validity, the Independence model would seem to be not only the most reliable and valid of the three models but also the one that provides substantially more reliable proportions than the raw (observed) data, particularly for small cells (e.g., in Table 3,  $r_{EE}$  for the Independence model in the case of the Retention-probability categories is .99 and  $r_{OO}$  is .42).

The superiority of the Independence model over the Linear-covariance model was unexpected since the latter model uses more information (i.e., the covariances) than the former. Some of this superiority may be attributable to differences in error variances. Each cell-frequency estimate by the Linear-covariance model is based on many covariance (i.e., error prone) terms, while the Independence model uses only one generalized variance term. The Independence model assumes that all of the covariances equal zero. However, it is perhaps arguable that some of the covariances do not equal zero. The covariance between the Naval Academy Commission Source and assignment to a small combatant ship on the first tour of duty, for example, must certainly be positive in the population of officers as a whole. Because the estimates are made separately in the retained and non-retained groups of officers, what seems to occur is that, within these two groups, the covariances do--as assumed by the Independence model--tend to equal zero. In terms of partial correlation, otherwise non-zero covariances approach zero when retention is partialled out. For the  $X_1$  Naval Academy Commission Source and the  $X_2$  Initial Assignment to a Small Combatant Ship, this explanation assumes a positive covariance between each of these variables and retention, which is indeed the case. The situation here thus seems to have the same structure as a common one involving three variables--height, age, and intelligence. Among children, intelligence has a high positive correlation with height, which is sharply reduced when age (positively correlated with each) is partialled out.

## CONCLUSIONS

1. The Structural Pattern Analysis (SPA) models investigated in the present study can provide stable, valid estimates of personnel retention proportions for possible use in the "cost" matrix of jobs and assignments to optimize allocation strategies.
2. Of the three SPA models evaluated in this study--(1) True-score, (2) Linear-covariance, and (3) Independence--the third one was the most accurate and stable, and it also provided more stable values than did the calculations based on the actual (observed) retention outcomes.
3. The SPA models, particularly the Independence model of the Covariance-structure type, are particularly useful for estimating retention proportions for patterns that contain few or no individuals.
4. Use of all available covariance terms in an SPA model appears to generate more error variance than true variance, particularly for patterns containing few individuals.
5. The data base used to test the SPA models in the present study was limited to individuals having two assignments within a specific experience range, whereas the complete data base includes many individuals who had only one assignment within this range. Research for further evaluation of SPA models would appropriately include (1) testing the models on a data base comprising a mix of individuals with one or two assignments and (2) comparing the results of an allocation strategy based on alternative inputs to the "cost" matrix from observed vs. SPA-generated values.

REFERENCES

- Gaier, E. L., & Lee, M. C. Pattern analysis: The configural approach to predictive measurement. Psychological Bulletin, 1953, 50, 141-149.
- Horst, P. Configural analysis and pattern recognition. Journal of Clinical Psychology Monograph, 1968, 25(4), 383-405.
- Lubin, A., & Osborn, H. G. A theory of pattern analysis for the prediction of a quantitative criterion. Psychometrika, 1957, 22, 63-73.
- Lubin, A., & Osborn, H. G. The use of configural analysis for the prediction of a qualitative criterion. Educational and Psychological Measurement, 1960, 20, 275-282.
- Lykken, D. T., & Rose, R. Psychological prediction from actuarial tables. Journal of Clinical Psychology, 1963, 19, 139-151.
- Meehl, P. E. Configural scoring. Journal of Consulting Psychology, 1950, 14, 165-171.
- Robertson, D. W., & Montague, W. E. Comparative racial analysis of Enlistment Advancement Exams: Relative item-difficulty between performance-matched groups (NPRDC Tech. Rep. 76-34). San Diego, CA: Navy Personnel Research and Development Center, March 1976.
- Robertson, D. W., & Pass, J. J. Relationship of officer first assignment and education major to retention (NPRDC Tech. Rep. 79-12). San Diego, CA: Navy Personnel Research and Development Center, March 1979.
- Solomon, H. Classification procedures based on dichotomous response vectors. In I. Clkin et al. (Eds.), Contributions to probability and statistics. Stanford, CA: Stanford University Press, 1960.
- Weitzman, R. A. Pattern scoring of short predictors. Proceedings of the 81st Annual Convention of the American Psychological Association, 1973, 27-28. (a)
- Weitzman, R. A. Pattern analysis: Method and application (Tech. Rep. NPS55WZ73091A). Monterey, CA: Naval Postgraduate School, 1973. (b)

**APPENDIX A**  
**OFFICER ASSIGNMENT CATEGORIES BY UNIT-TYPE**

OFFICER ASSIGNMENT CATEGORIES BY UNIT-TYPE

Unit Category			Ship and Station Code (SSC)
No.	Abbreviation	Title	Sources <sup>a</sup>
01	AIR-SQD/GP	Air-Squadron/Staff/Group	05, 08A, 09 (except AHK), 11, 14, 15
02	CVAN	Carrier-Nuclear Propulsion	10C
03	CV	Carrier (all except nuclear)	10ABDEFGZ
04	AMPHIB	Amphibious (except LST)	17 (except M), 18
05	LST	Tank Landing Ship	17M
06	CA/CL/BB	Cruisers (except Guided Missile) and Battleship	19, 21AZ, 22ABCZ
07	CG	Cruiser (Guided Missile)	21BCD, 22D
08	DD/DL	Destroyer (except Guided Missile and Radar)	23ABFZ
09	DD/DE-RAD	Destroyer (Radar)	23EC, 24C
10	DD/DL/DE-GUID	Destroyer (Guided Missile)	23DGH, 24D
11	DE	Destroyer Escort (except Radar)	24ABZ
12	STAFF-JT/FLT	Staff-Joint/Fleet	08N, 09EK, 61E, 64
13	TEND-REP	Tender (except Destroyer) Repair	36, 41, 47, 48, 49, 50, 88
14	AE	Ammunition	16
15	AF/AK/AV	Cargo	20
16	AD	Destroyer Tender	39
17	AP/AH	Transport	28, 51, 52
18	MNSWP	Minesweeper	32, 33, 34
19	STAFF-AMPH, FMF	Staff-Amphibious and Fleet Marine	71 (except EF), 72
20	COMM-SECUR	Communications and Security	86
21	INTELL	Intelligence	76
22	DIPLOM	Diplomatic	66, 6-A, 6-E
23	OCEANOG	Oceanographic	69
24	AUX/MERCH	Auxiliary and Merchant	25, 35
25	TUG-O	Tug-Ocean	53
26	PC-GUN	Gunboat	27, 37, 40, 45, 46
27	MNLAY	Mine Warfare	29, 30, 31, 38, 75
28	CB-SHIPYD	Construction	67, 71E, 81, 99
29	RESC-SALV	Rescue-Salvage	42, 43

Note. Reproduced from Robertson and Pass (1979), Table 2.

<sup>a</sup>SSCs are defined in the Officer Classification Manual, NAVPERS 15839C, Volume I, Part H. Definitions of the SSCs were reproduced in Robertson and Pass (1979).

## OFFICER ASSIGNMENT CATEGORIES BY UNIT-TYPE (Continued)

Unit Category			Ship and Station Code (SSC)
No.	Abbreviation	Title	Sources <sup>a</sup>
30	ADVBASE	Advanced Base	54, 55
31	BASE-DEPOT	Bases and Depots	62, 65, 79, 87, 90
32	AMMO DEP	Ammunitions Depot	60
33	ORD RANGE	Ordnance Ranges	84
34	ED-TRA	Education and Training	08X, 91, 97, 98
35	R&D	Research and Development	89
36	SYSCOM	Systems Command	56, 58, 70, 83, 85, 92, 93, 94, 95, 96
37	JT ACT	Army/Navy/Air Force Joint Activities	61 (except E)
38	GOVT AGENCY	Government Agencies	68
39	PERS	Personnel Activities	77, 78
40	NAV-DEPT/OP	Navy Department and Operations	80, 82
41	STAFF-F	Staff-Force	08EFHKMRTVY, 09A, 71F
42	STAFF-G(NA)	Staff-Group (Non-Air)	08CDGJLPQSUWZ
43	AIR-STA/TRA	Air-Station/Training	08B, 57, 59

Note. Reproduced from Robertson and Pass (1979), Table 2.

<sup>a</sup>SSCs are defined in the Officer Classification Manual, NAVPERS 15839C, Volume I, Part H. Definitions of the SSCs were reproduced in Robertson and Pass (1979).

**APPENDIX B**  
**DESCRIPTION OF STRUCTURAL PATTERN ANALYSIS MODELS**

## DESCRIPTION OF STRUCTURAL PATTERN ANALYSIS MODELS

Each of the three models to be described below provides an estimate of the proportion of individuals in a cell defined by the values of two or more variables (e.g., Commission Source and First and Second Duty Assignment). Combination of this proportion for retained ( $p_R$ ) and nonretained ( $p_N$ ) groups of officers results in a Retention-probability estimate ( $P_R'$ ) for the cell:

$$P_R' = \frac{n_R p_R}{n_R p_R + n_N p_N},$$

where  $n_R$  is the total number of retained and  $n_N$  the total number of nonretained officers in the sample. This formula is analogous to the Bayes formula used to determine posterior probabilities.

### True-score Model

The True-score model (derived by the first author in a separate report under preparation) uses the linear regression of true proportions on observed proportions to obtain for each observed proportion ( $X$ ) a true-proportion estimate of  $X$ , designated  $X'$ :

$$X' = r_{XX} X + (1 - r_{XX})/N,$$

where  $N$  is the total number of patterns and  $r_{XX}$  is an estimate of the reliability of  $X$ . Since  $r_{XX}$  must be between zero and one,  $X'$  will tend to be between  $X$  (when  $r_{XX} = 1$ ) and  $1/N$  (when  $r_{XX} = 0$ ). If  $X = 0$ , in particular, then  $X'$  can be no smaller than zero. This is the principal advantage of the model: It never yields estimates less than zero. Indeed, whenever  $r_{XX}$  is less than one, which is the usual case, all estimates must be greater than zero.

The reliability estimator used in this study is

$$r_{XX} = 1 - N(1 - \sum X^2)/(M - 1)(N \sum X^2 - 1),$$

where  $\sum X^2$  is the sum of the squares of the  $N$  observed proportions (one for each pattern) computed from the sample of  $M$  individuals.

### Linear-covariance Model

Covariance-structure estimation, described by Solomon (1960), makes use of a generalization to three or more variables of the standard covariance formula for 0-1 binary variables,  $X_1$  and  $X_2$ :

$$\text{Cov}(X_1, X_2) = p_{12} - p_1 p_2,$$

where  $p_{12}$  is the proportion of individuals for whom both  $X_1 = 1$  and  $X_2 = 1$  and  $p_i$  is the proportion of individuals for whom  $X_i = 1$  ( $i = 1, 2$ ). If  $\text{Cov}(X_1, X_2) \approx 0$ , that is, if  $X_1$  and  $X_2$  tend to be independent, then  $p_{12} \approx p_1 p_2$  so that the product  $p_1 p_2$  is an estimator of the cell (pattern) proportion  $p_{12}$ .

The generalization to three binary variables  $X_1$ ,  $X_2$ , and  $X_3$  (as in the present study) is

$$p_{123} \approx p_1 p_2 p_3 + p_1 \text{Cov}(X_2, X_3) + p_2 \text{Cov}(X_1, X_3) + p_3 \text{Cov}(X_1, X_2).$$

The second model investigated, the Linear-covariance model, uses this three-variable approximation to estimate the cell proportion  $p_{123}$ . The  $X$ 's in the True-score model thus correspond to the  $p_{123}$  values here; the  $X$ 's here have a different meaning:  $X_1 = 1$  for an individual who has a specific commission source ( $X_1 = 0$  otherwise) and  $X_i = 1$  ( $i = 2, 3$ ) for an individual whose  $(i - 1)$ th duty assignment is to a specific billet category ( $X_i = 0$  otherwise).

Because covariances can be less than zero, this approximation can yield negative values of  $p_{123}$ . The Linear-covariance model thus requires rescaling of the estimates to avoid values less than zero. The rescaling equation used was linear--hence the use of the word linear in the name of the model--and the determination of its constants satisfied two conditions:

1. The mean estimated proportion had to be equal to the mean observed proportion.
2. The sum of the smallest estimates for the retention and nonretention groups of officers had to be rescaled to zero.

Condition 2 ensured that all rescaled estimates would be larger than zero.

#### Independence Model

The third model is a special case of the second in which all covariances are assumed to be approximately equal to zero:

$$p_{123} \approx p_1 p_2 p_3$$

Since this approximation assumes that  $X_1$ ,  $X_2$ , and  $X_3$  tend to be independent, the model is called the Independence model. Unlike the simple covariance model (without rescaling), this model cannot yield estimates less than zero. The estimates of this model are also unbiased, since their mean is algebraically equal to the mean of the observed proportions.

APPENDIX C

RETENTION ESTIMATES BY THE  
SPA INDEPENDENCE MODEL

Table C-1

Comparison of Observed Retention Proportions with  
 Estimates by the SPA Independence Model  
 for 24 High-retention Patterns of  
 the Ship-type Categories

Assignment pattern $X_1 X_2 X_3^a$	Retention proportion					
	N <sup>b</sup>	Subgroup 1		Subgroup 2		
		Estimated $E_1$	Observed $O_1$	N <sup>b</sup>	Estimated $E_2$	Observed $O_2$
111	146	.93	.83	146	.92	.75
131	17	.85	.71	17	.82	.65
121	10	.83	.90	11	.82	.73
113	21	.81	.62	20	.79	.70
141	1	.80	1.00	1	.74	1.00
211	106	.79	.62	106	.73	.42
114	15	.78	.67	15	.73	.53
115	59	.75	.58	59	.75	.66
112	14	.74	.57	14	.65	.43
151	1	.72	1.00	0	.67	--
161	1	.70	.00	0	.66	--
116	55	.69	.62	55	.69	.66
511	33	.65	.39	33	.52	.52
133	4	.64	1.00	4	.61	.75
123	0	.62	--	0	.60	--
231	27	.62	.52	28	.53	.61
134	3	.60	.68	3	.54	.67
221	37	.59	.43	37	.52	.49
311	56	.58	.36	55	.49	.26
124	2	.58	.50	2	.53	.50
143	0	.57	--	0	.49	--
135	8	.56	.25	8	.56	.50
213	14	.56	.57	13	.48	.23
132	2	.54	.00	1	.44	.00

<sup>a</sup>See page 3 for Commission Source ( $X_1$ ) code. See Table 1 for Ship-type categories ( $X_2$ --first assignment;  $X_3$ --second assignment).

<sup>b</sup>Total number of officers assigned to the pattern.

Table C-2

Comparison of Observed Retention Proportions with  
 Estimates by the SPA Independence Model  
 for 24 Low-retention Patterns of  
 the Ship-type Categories

Assignment pattern $x_1 x_2 x_3^a$	Retention proportion					
	Subgroup 1			Subgroup 2		
	N <sup>b</sup>	Estimated E <sub>1</sub>	Observed O <sub>1</sub>	N <sup>b</sup>	Estimated E <sub>2</sub>	Observed O <sub>2</sub>
466	264	.02	.04	265	.02	.04
456	65	.02	.05	64	.02	.06
462	6	.02	.17	6	.01	.17
465	33	.02	.00	33	.02	.03
452	5	.03	.20	6	.01	.00
455	46	.03	.09	47	.02	.02
464	9	.03	.11	9	.02	.11
446	148	.03	.05	149	.02	.05
454	9	.03	.00	9	.02	.00
463	2	.03	.00	3	.03	.00
426	128	.04	.06	127	.03	.04
453	6	.04	.00	5	.03	.00
366	20	.04	.10	19	.03	.21
442	18	.04	.00	17	.02	.00
436	142	.04	.08	143	.04	.05
445	69	.04	.07	70	.03	.04
356	4	.04	.00	5	.03	.20
422	28	.05	.11	29	.03	.03
425	70	.05	.04	70	.05	.03
444	92	.05	.09	92	.03	.07
362	1	.05	.00	0	.03	--
566	45	.05	.13	44	.04	.05
432	22	.05	.05	23	.03	.00
365	1	.05	.00	2	.04	.00

<sup>a</sup>See page 3 for Commission Source ( $x_1$ ) code. See Table 1 for Ship-type categories ( $x_2$ --first assignment;  $x_3$ --second assignment).

<sup>b</sup>Total number of officers assigned to the pattern.

Table C-3

Comparison of Observed Retention Proportions with Estimates  
by the SPA Independence Model for 24 High-retention  
Patterns of the Retention-probability Categories

Assignment pattern	$X_1 X_2 X_3^a$	Retention proportion					
		Subgroup 1			Subgroup 2		
		N <sup>b</sup>	Estimated E <sub>1</sub>	Observed O <sub>1</sub>	N <sup>b</sup>	Estimated E <sub>2</sub>	Observed O <sub>2</sub>
111	36	.95	.75	35	.94	.74	
112	15	.95	.93	14	.93	.64	
121	50	.92	.82	49	.91	.78	
122	46	.91	.80	47	.91	.81	
131	15	.90	.60	14	.86	.79	
113	7	.90	.86	8	.88	.75	
132	5	.90	.60	6	.85	.83	
211	27	.85	.67	27	.78	.63	
212	7	.84	.57	8	.77	.50	
123	11	.84	.55	12	.84	.83	
115	3	.84	1.00	4	.80	.75	
141	2	.82	.50	2	.79	1.00	
142	1	.82	.00	2	.78	1.00	
151	2	.82	1.00	2	.79	1.00	
133	3	.81	1.00	3	.75	.33	
117	18	.81	.89	18	.80	.44	
152	2	.81	1.00	1	.78	1.00	
114	1	.79	.00	2	.80	.00	
161	1	.78	.00	2	.73	1.00	
162	0	.77	--	0	.72	--	
221	48	.77	.50	48	.72	.38	
116	3	.76	.33	3	.73	.00	
222	24	.76	.63	23	.71	.44	
118	24	.75	.67	24	.75	.71	

<sup>a</sup>See page 3 for Commission Source ( $X_1$ ) code. See Table 1 for Retention-probability categories ( $X_2$ --first assignment;  $X_3$ --second assignment).

<sup>b</sup>Total number of officers assigned to the pattern.

Table C-4

Comparison of Observed Retention Proportions with Estimates  
 by the SPA Independence Model for 24 Low-retention  
 Patterns of the Retention-probability Categories

Assignment pattern $X_1 X_2 X_3$	Retention proportion					
	Subgroup 1			Subgroup 2		
	N <sup>b</sup>	Estimated E <sub>1</sub>	Observed O <sub>1</sub>	N <sup>b</sup>	Estimated E <sub>2</sub>	Observed O <sub>2</sub>
488	264	.02	.04	265	.02	.05
486	9	.02	.11	9	.02	.11
478	65	.02	.06	64	.02	.05
484	1	.02	.00	0	.02	--
476	8	.02	.00	8	.02	.00
487	33	.02	.00	33	.02	.03
474	1	.02	.00	2	.02	.00
468	157	.02	.06	156	.02	.04
485	4	.03	.00	5	.02	.20
477	47	.03	.04	46	.02	.07
466	46	.03	.04	45	.02	.04
475	5	.03	.00	5	.02	.00
458	94	.03	.03	93	.03	.07
464	12	.03	.17	11	.03	.00
456	33	.03	.03	34	.03	.06
448	101	.03	.09	101	.03	.07
388	20	.03	.15	19	.04	.16
446	35	.03	.03	35	.03	.09
467	72	.03	.00	72	.03	.03
386	0	.03	--	1	.04	.00
454	12	.04	.00	12	.04	.25
588	44	.04	.07	45	.05	.11
378	4	.04	.25	5	.04	.00
586	2	.04	.50	2	.05	.00

<sup>a</sup>See page 3 for Commission Source ( $X_1$ ) code. See Table 1 for Retention-probability categories ( $X_2$ --first assignment;  $X_3$ --second assignment).

<sup>b</sup>Total number of officers assigned to the pattern.

## DISTRIBUTION LIST

No. of copies

Defense Documentation Center Cameron Station, Bldg. 5 5010 Duke Street Alexandria, VA 22314	2
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	1
Library (Code 0142) Naval Postgraduate School Monterey, CA 93940	2
Library (Code 54) Naval Postgraduate School Monterey, CA 93940	4
Professor J. K. Arima Professor R. S. Elster Professor C. R. Jones Professor D. M. Rousseau Professor J. D. Senger Professor R. A. Weitzman Code 54 Naval Postgraduate School Monterey, CA 93940	1 1 1 1 1 34
Dr. D. W. Robertson Code 310 Navy Personnel Research & Development Center San Diego, CA 92152	34
Principal Deputy Assistant Secretary of the Navy (Manpower and Reserve Affairs)	1
Chief of Naval Operations (OP-102) (OP-11) (OP-987H)	1 1 1
Chief of Naval Research (Code 450)	4
Chief of Information (OI-2252)	1
Director of Navy Laboratories	1
Officer in Charge, Navy Occupational Development and Analysis Center	2
Director, Training Analysis & Evaluation Group	1
Personnel Research Division, Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base	1

Occupational & Manpower Research Division Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base	1
Technical Library Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base	1
Program Manager Life Sciences Directorate Air Force Office of Scientific Research (AFSC)	1
Army Research Institute for the Behavioral & Social Sciences	1
Military Assistant for Training & Personnel Technology Office of the Under Secretary of Defense for Research & Engineering	1
Director of Acquisition Planning Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics	1
Library Operations Section Library of Congress	1
Library Navy Personnel Research and Development Center San Diego, CA 92152	3